

高光谱反射率的滨海地区土壤全磷含量反演

魏丹萍, 郑光辉*

南京信息工程大学地理科学学院, 江苏 南京 210044

摘要 反射光谱在近年来广泛应用于土壤属性的估算。作为一种有效估算土壤全磷含量的手段, 反射光谱技术可以很大程度上减少传统化学测量方法所损耗的人力物力。以江苏滨海土壤为研究对象, 在30个采样点采集了共147个土样, 测量土壤样品光谱反射率及全磷含量。利用原始光谱反射率数据及6种不同的光谱变换结果, 通过随机抽样(RS)、KS、SPXY三种样本集划分方法, 基于偏最小二乘回归(PLSR)和支持向量机(SVM)方法分别建立土壤全磷含量的估算模型, 对比分析了三种样本集划分方法对估算结果精度的影响。结果表明: (1)以原始光谱反射率为数据, PLSR模型, RS方法在多数情况下可以获得较为稳定的模型精度, 明显优于KS和SPXY方法; 在SVM模型中, 采用SPXY方法获得的模型结果最优, KS次之, RS结果最差。(2)不同的样本集划分方法所合适的光谱变换方法不同, 对于三种划分样本集方法, PLSR和SVM对应的最优光谱变换分别是对数的倒数和一阶导数(KS方法), 原始光谱和一阶导数(RS方法), 一阶导数和多元散射校正(SPXY方法)。其中采用KS方法划分样本集, PLSR和SVM均能获得最佳的预测结果。并非所有光谱变换方法都可以提高模型精度, 部分光谱变换后PLSR模型预测精度显著降低; (3)在所有的样本集划分方法中, SVM的建模效果优于PLSR, 采用RS方法划分样本集, PLSR的预测精度高于SVM, 而采用KS和SPXY方法划分样本集, SVM的预测精度整体高于PLSR。综上所述, 本研究区域估算土壤全磷含量的最佳模型是基于KS样本集划分方法和一阶导数光谱变换建立的SVM模型, 此时拟合优度(R_p^2)为0.82。结果表明反射光谱可以对滨海地区的土壤全磷含量进行有效预测, 对土壤磷元素的高效快速反演具有一定的指导意义。

关键词 全磷; 反射光谱; 光谱变换; 样本划分方法; 偏最小二乘回归; 支持向量机

中图分类号: S153.6 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2022)02-0517-07

引言

作为土壤中仅次于氮的营养成分, 磷在很大程度上影响着土壤肥力的高低, 土壤磷的流失也是导致湖泊富营养化的主要因素, 因此快速准确获取土壤全磷含量对于农业生产和环境保护具有重要意义。传统的化学测量方法需要耗费大量人力物力, 且不能大面积的推广, 给土壤全磷含量的估算带来了一定的困难。近年来, 土壤反射光谱被广泛的应用于土壤属性估算, 由于其获取方式容易, 价格低廉等特点, 已成为估算土壤属性的重要手段。利用反射光谱可以预测不同土壤类型的全磷含量, 张佳佳等^[1]发现使用反射光谱可以较好地预测南方丘陵红壤的土壤全磷含量; 薛利红等^[2]同样利用反射光谱成功反演了太湖流域水稻土的全磷含量。不同土壤光

谱数据的最佳光谱变换方法和预测模型也可能不同, Xiong等^[3]将倒数光谱变换结合多元逐步回归方法预测巢湖地区土壤磷含量; 王莉雯等^[4]将原始数据结合偏最小二乘回归方法建立湿地土壤的全磷含量模型; Sarathjith等^[5]将一阶导数变换结合支持向量机方法估算土壤全磷含量。此外, 不同的样本集划分方式会对建模结果产生影响, 陈奕云等^[6]采用Kennard-Stone(KS)和Sample Set Partitioning Based on Joint X-Y Distance Algorithm(SPXY)方法划分样本集研究江汉平原洪湖地区水稻土中土壤有机质含量; 于慧春等^[7]利用SPXY方法划分样本建立霉变玉米中两种毒素含量的预测模型; 毛博慧^[8]等比较了随机抽样(random sampling, RS)和SPXY样本集划分方法对于冬小麦叶绿素浓度的估算精度影响。全磷含量的估算也可能受其他因素的影响, 如章文龙等^[9]发现闽江河口湿地土壤全磷与有机质诊断指数高度

收稿日期: 2020-12-29, 修订日期: 2021-01-28

基金项目: 国家自然科学基金项目(41201215, 41877004)资助

作者简介: 魏丹萍, 女, 1997年生, 南京信息工程大学地理科学学院硕士研究生 e-mail: wdpnuist@163.com

* 通讯作者 e-mail: zgh@nuist.edu.cn

相关。

目前我国对于土壤中全磷含量的光谱特征研究仍处于初级阶段,不同土壤类型之间的属性估算模型存在明显差异。滨海滩涂土壤作为我国重要的后备土地资源 and 空间资源,面积广且分布集中,具有较大的开发潜力。因此,对滨海土壤的光谱特征及属性反演研究具有重要的科学与实践意义。为了提高模型精度,很多学者采用不同的光谱变换方法来处理光谱数据,但较少将不同的样本集划分方法应用于土壤全磷含量的估算研究中,或将不同的光谱变换结合数据集划分方法来建立模型并分析其对估算精度的影响。因此,为了探索不同样本集划分方法以及不同光谱变换结合数据集划分方法对模型精度的影响,我们利用光谱反射率数据结合偏最小二乘回归和支持向量机两种建模方法开展以下研究:(1)对比分析三种样本集划分方法对土壤全磷含量估算结果的影响;(2)评价不同的光谱变换方法结合不同样本数据集划分方法对于滨海土壤全磷含量模型精度的影响;(3)评估利用反射光谱估算滨海地区土壤全磷含量的可靠性。

1 实验部分

1.1 样本采集

东台市位于江苏省沿海地区,长江三角洲北翼,境内地势平坦,河流众多。该地区属于亚热带季风海洋性气候,年平均气温为 15.0℃,年降水量为 1 061.2 mm。东台地区的土壤类型以水稻土和潮盐土为主,两者大致以范公堤为界,境内的滩涂资源十分丰富。本研究在东台地区采集 13 个 0~5 cm 的表土样点和 17 个 0~100 cm 的剖面样点,共计 30 个采样点(图 1)。剖面采样深度分别为 0~5, 5~10, 10~20, 20~30, 30~40, 40~60, 60~80 和 80~100 cm,由于两个剖面分别缺失一个 0~5 和 80~100 cm 的样本,最终获得 147 个有效样品。将土样风干并研磨,过 100 目筛,分成两份分别用于光谱测量和化学分析,采用硫酸-高氯酸消化-钼锑抗比色法测量土壤全磷含量,利用重铬酸钾氧化-外加热法测量土壤有机质含量。



图 1 采样点位置分布图
Fig. 1 Sampling location

1.2 光谱测量

土壤的光谱测量在暗室中进行,采用美国分析光谱仪器公司(ASD)的 FieldSpec3 便携式光谱分析仪,该设备的探测波段在 350~2 500 nm 之间。实验室的环境条件设置如下:

将一盏 50 W 的卤素灯作为光源,调整光照角度为 15°,距离样品 30 cm;采用 5°视场探头角,探头在距土样 15 cm 处进行光谱采集,探测半径为 0.655 cm。在测量光谱之前,先去暗电流,利用空白标准参照板获取相对反射率,取适量土样放入器皿内并将表面刮平,土样放入平台匀速转动,同时可获得 20 条光谱曲线,算数平均后得到最终的反射光谱数据。

1.3 光谱预处理及光谱变换

光谱数据在 1 000 和 1 800 nm 处产生明显的断点,利用仪器自带的软件 ViewSpecPro 对断点进行校正,获得较为连续的光谱曲线。利用 Savitzky-Golay(SG)方法将光谱曲线作平滑处理,去除前后 10 nm 波段的光谱数据,保留 360~2 490 nm 波段范围的光谱曲线作为原始光谱反射率(reflectivity, R)。对 R 分别进行标准正态变量变换(standard normal variable, SNV)、多元散射校正(multiple scatter correction, MSC)、一阶微分(first derivative, FD)、二阶微分(second derivative, SD)、对数的倒数(reciprocal of logarithm, RL)和连续统去除(continuum removal, CR)等共 6 种不同的光谱变换。SNV 方法主要是用来消除固体颗粒大小、表面散射以及光程变化对近红外漫反射光谱的影响。MSC 方法可以有效地消除散射影响,增强成分含量相关的光谱吸收信息,提高信噪比。光谱数据经过 FD 处理后特征差异明显增大,线性背景和噪声等干扰降低。SD 光谱变换可以很大程度上消除原始光谱的背景干扰,提高分析精度^[10]。RL 方法可以增强光谱的差异,并减少光照条件变化对于光谱的影响。CR 方法主要将反射率归一化为 0~1,从而提取特征波段以便识别^[11]。

1.4 样本集划分方法

除了最常用的随机抽样(RS)方法,还采用了 KS 和 SPXY 样本集划分方法。为减小随机划分样本集对模型结果的影响,进行 1 000 次随机抽样、建模及预测,综合分析其 1 000 次的计算结果。KS^[12]方法是根据样本自变量之间的欧氏距离来选择建模集和预测集。SPXY 方法的基本思路与 KS 方法相似,最大改变在于它还考虑了因变量 Y 之间的欧氏距离,将自变量 X 和因变量 Y 之间的欧氏距离分别进行最大值标准化后求和,以此“和”的值作为样本之间的距离,再按照 KS 方法的迭代思路建立建模集。建模集样本数量占 80%(118 个),预测集样本数量占 20%(29 个)。

1.5 建模方法和模型评价

偏最小二乘回归^[13]是一种集主成分分析、典型相关分析和多元线性分析方法优点于一体的新型多变量回归分析方法,能够简化数据结构,并解决自变量的多重相关性问题。支持向量机^[14]在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势,并能推广应用到函数拟合等其他问题中。

偏最小二乘回归(partial least square regression, PLSR)和支持向量机(support vector machine, SVM)的模型精度取决于拟合优度(R^2)、均方根误差(root mean square error, RMSE)、相对分析误差(relative percent deviation, RPD)三个评价指标^[15]。RMSE(建模集 RMSE_c、预测集 RMSE_p)趋

近于 0、 R^2 (建模 R_c^2 、预测 R_p^2) 趋近于 1、RPD 越大, 说明模型的稳定性越好、拟合程度越高。

2 结果与讨论

2.1 土壤属性特征

使用标准差、变异系数、峰度和偏度等描述土壤属性特征。变异系数 (variable coefficient, CV) 是指数据的离散程度; 峰度 (kurtosis, K) 可以描述数据分布形态的陡缓程度, 3 代表正态分布, 1.8 代表均匀分布; 偏度 (skewness, S) 可以衡量数据分布的偏斜程度, 0 代表正态分布。由表 1 可见, 东台地区滩涂土壤有机质 (soil organic matter, SOM) 含量较低 (平均值为 $9.18 \text{ g} \cdot \text{kg}^{-1}$), 具有很强的空间异质性 (最小值

为 $1.39 \text{ g} \cdot \text{kg}^{-1}$, 最大值为 $52.56 \text{ g} \cdot \text{kg}^{-1}$, CV 为 95.21%)。全磷 (total phosphorus, TP) 含量在 $0.22 \sim 1.30 \text{ g} \cdot \text{kg}^{-1}$ 之间, 其变化幅度低于 SOM 含量变化, CV 仅为 28.02%。SOM 的 $K > 3$ ($K = 6.72$), 表明数据分布中具有过高的峰值。TP 的 $K < 3$ ($K = 1.01$) 则代表数据分布中具有过低的峰值, 由此可以看出 TP 相对于 SOM 分布比较平坦。TP 和 SOM 的偏度均为正值, 呈正偏态分布, 数据的平均值大于中位数和众数, 表示样本中存在部分 TP 和 SOM 含量过高的点。而 TP 的偏度绝对值高于 SOM, 代表其数据的偏倚程度较大。在本研究区中 TP 与 SOM 之间相关性弱 (相关系数为 0.17), 标准差、变异系数、峰度和偏度相差较大, 因此无法利用 SOM 来研究 TP 对土壤光谱的影响。

表 1 土壤 TP 与 SOM 含量

Table 1 Statistics of soil TP and SOM contents

土壤属性	最大值/ ($\text{g} \cdot \text{kg}^{-1}$)	最小值/ ($\text{g} \cdot \text{kg}^{-1}$)	平均值/ ($\text{g} \cdot \text{kg}^{-1}$)	标准差/ ($\text{g} \cdot \text{kg}^{-1}$)	变异系数/ %	峰值 K	偏度 S
TP	1.30	0.22	0.68	0.19	28.02	1.01	2.19
SOM	52.56	1.39	9.18	8.74	95.21	6.72	0.61

2.2 样本集划分

KS 和 SPXY 算法在样本集的划分中应用广泛, 可以在一定程度上提高模型的预测精度。利用 KS 和 SPXY 算法对样品集进行划分, 将土壤全磷样本划分成为 118 个建模集和 29 个预测集, KS 和 SPXY 样本划分的结果见表 2 和表 3。从两个表中可以看出, 建模集样本的 TP 含量涵盖了预测集样

本的 TP 含量范围, 变异系数小于 31%, 样本集划分较合理。KS 划分样本的建模集变异系数略低于 SPXY 方法, 而其预测集的变异系数始终高于 SPXY 方法。利用同一种方法划分样本, 不同的光谱变换获取的建模和预测样本集也略有不同, 因此得到的建模结果可能有所差异。

表 2 KS 样本集划分

Table 2 Sample set partitioning based on KS

统计参数	R		SNV		MSC		FD		SD		RL		CR	
	建模集	预测集	建模集	预测集	建模集	预测集	建模集	预测集	建模集	预测集	建模集	预测集	建模集	预测集
最小值/ $(\text{g} \cdot \text{kg}^{-1})$	0.22	0.46	0.22	0.51	0.22	0.51	0.22	0.47	0.22	0.44	0.22	0.50	0.22	0.46
最大值/ $(\text{g} \cdot \text{kg}^{-1})$	1.30	1.28	1.30	1.14	1.30	1.14	1.30	1.14	1.30	1.14	1.30	1.28	1.30	1.14
平均值/ $(\text{g} \cdot \text{kg}^{-1})$	0.67	0.76	0.68	0.72	0.68	0.72	0.68	0.72	0.69	0.67	0.67	0.74	0.68	0.71
标准差/ $(\text{g} \cdot \text{kg}^{-1})$	0.19	0.17	0.20	0.13	0.20	0.13	0.20	0.13	0.20	0.17	0.20	0.17	0.20	0.16
变异系数/%	29.11	22.19	30.26	17.99	30.26	17.99	30.16	18.63	28.83	25.15	29.33	22.36	29.48	22.43

表 3 SPXY 样本划分

Table 3 Sample set partitioning based on SPXP

统计参数	R		SNV		MSC		FD		SD		RL		CR	
	建模集	预测集	建模集	预测集	建模集	预测集	建模集	预测集	建模集	预测集	建模集	预测集	建模集	预测集
最小值/ $(\text{g} \cdot \text{kg}^{-1})$	0.22	0.49	0.22	0.45	0.22	0.45	0.22	0.50	0.22	0.47	0.22	0.49	0.22	0.46
最大值/ $(\text{g} \cdot \text{kg}^{-1})$	1.30	0.84	1.30	0.84	1.30	0.83	1.30	0.83	1.30	0.84	1.30	0.84	1.30	1.08
平均值/ $(\text{g} \cdot \text{kg}^{-1})$	0.69	0.68	0.68	0.70	0.68	0.70	0.68	0.72	0.69	0.66	0.69	0.68	0.68	0.69
标准差/ $(\text{g} \cdot \text{kg}^{-1})$	0.21	0.13	0.21	0.11	0.21	0.11	0.21	0.10	0.21	0.12	0.21	0.12	0.20	0.14
变异系数/%	30.01	18.61	30.58	15.57	30.57	15.67	30.89	13.50	29.91	17.82	30.09	18.07	29.83	20.20

2.3 各数据集划分方法的建模、预测验证结果分析

表 4 和表 5 分别是利用原始光谱数据建立的基于三种数据集划分方法的 PLSR 和 SVM 预测模型。从 PLSR 的结果

可知, RS 方法的 1000 次运算结果差异较大, 除去预测 R_p^2 小于 0 的异常次数 (9 次), 建模 R_c^2 范围为 $0.29 \sim 0.78$, 预测 R_p^2 范围为 $0.01 \sim 0.82$, 说明 RS 方法每次抽取的建模数据集

不同, 所得到的结果精度也不相同, 且差异较大; R_c^2 平均值为 0.67, R_p^2 平均值为 0.51, RPD 平均值为 1.49, 潜变量 (latent variable, LV) 的平均值为 9.14; 在 991 次运算结果中, 有 97.38% (965 次) 的 R_c^2 和 58.22% (577 次) 的 R_p^2 大于 0.50; 57.92% (574 次) 的结果满足 R_c^2 与 R_p^2 同时大于 0.50。另外, 在正常情况下, 合理的建模及验证结果也应符合 R_c^2 大于 R_p^2 的一般规律。图 2 第一象限中 457 次位于对角线以下, 符合 R_c^2 大于 R_p^2 条件。所以在 1 000 次结果中, 利用 RS 方法划分的数据集结合 PLSR, 45.70% 可能性得到合理且精度较高的土壤全磷含量估算结果。图 2 显示 RS 的大部分估算结果和部分建模结果 (755 次的 $R_p^2 > 0.42$, 298 次的 $R_c^2 > 0.69$) 优于 KS 和 SPXY。将 RS-Mean 方法的结果和 KS 及 SPXY 方法对比, 发现三种方法的建模精度相差较小, 但 SPXY 的 R_c^2 略高于其余两种方法, 可视作建模效果最好。从预测效果来看, RS-Mean 方法的 R_p^2 和 RPD 均有最大值, 且 R_p^2 分别高于 KS 和 SPXY 约 0.1 和 0.25, RPD 则高了 0.15 和 0.3。因此, 即使 SPXY 的 $RMSEP_p$ 略小于其他两种方法, 也被认为总体精度较小。综合 PLSR 的结果来看, 利用原始反射光谱, RS-Mean 方法结果最优, KS 次之, SPXY 结果最差。

表 4 偏最小二乘回归结果

Partitioning methods	RMSE _c	R_c^2	RMSE _p	R_p^2	RPD	LVs
Min	0.09	0.29	0.07	0.01	1.02	3
RS Max	0.16	0.78	0.19	0.82	2.40	13
Mean	0.11	0.67	0.13	0.51	1.49	9.14
KS	0.11	0.67	0.13	0.42	1.34	9
SPXY	0.11	0.69	0.11	0.26	1.19	9

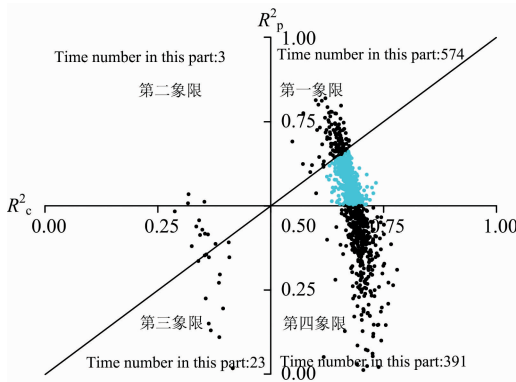
图 2 PLSR 建模及预测验证 R^2 散点分布图

Fig. 2 Scatter plot of determination coefficients for partial least square regression calibration and prediction

和 PLSR 的结果不同, 采用 SVM 方法建模, RS 划分样本集的 1 000 次运算结果差异相对较小, 建模 R_c^2 范围为 0.73~0.92, 预测 R_p^2 范围为 0~0.61, R_c^2 平均值为 0.84, R_p^2 平均值为 0.20, RPD 平均值为 1.11。在 1 000 次运算结

果中, 所有次数的 R_c^2 都大于 0.5 且高于 R_p^2 , 但只有 0.12% (12 次) 的 R_p^2 大于 0.5, 表明 SVM 方法的建模精度远高于预测精度。将 RS-Mean 方法的结果和 KS 及 SPXY 方法对比, 发现三种方法的建模精度相差较小, RS-Mean 方法的 R_c^2 略高于其他两种方法, 但其预测精度较低, R_p^2 仅为 0.2。预测精度最高的模型是采用 SPXY 方法划分样本集, R_p^2 达 0.46, 其 RPD 也高于其他两种方法约 0.27 和 0.29。综合 SVM 的结果, 原始反射光谱的 SPXY 方法结果最优, KS 次之, RS-Mean 结果最差。这与 PLSR 获得的结果相反, 表明不同的样本集划分方法会对建模预测结果产生较大的影响, 不同的建模方法适合的样本集划分方法也有所差异。

表 5 支持向量机的结果

Table 5 Statistics of support vector machine

Partitioning methods	RMSE _c	R_c^2	RMSE _p	R_p^2	RPD
Min	0.06	0.73	0.10	0.00	0.86
RS Max	0.11	0.92	0.27	0.61	1.42
Mean	0.09	0.84	0.17	0.20	1.11
KS	0.11	0.80	0.16	0.41	1.09
SPXY	0.10	0.83	0.09	0.46	1.38

2.4 模型精度分析

将 360~2 490 nm 波段的原始光谱数据, 分别经过 6 种光谱变换方法, 3 种数据集划分方法, 结合偏最小二乘回归方法和支持向量机方法分别建立模型, 并通过精度比较确定最优估算模型。从表 6 的 PLSR 分析结果可得: (1) 采用 RS 数据集划分方法, SD 光谱变换方法的建模精度最高, R_c^2 为 0.68, 但它的预测精度最低, R_p^2 仅为 0.44。综合来说, 该样本集划分方法的最佳模型是 SG 平滑后原始光谱数据, 它的估算精度 R_p^2 达到 0.50, 建模精度 R_c^2 也达到了 0.67。RL 和 MSC 光谱变换方法次之, 这两种方法的 R_p^2 均达到 0.48, 对应的 R_c^2 分别是 0.67 和 0.66; 其余光谱变换方法获得的结果也较好。(2) 采用 KS 样本集划分方法, FD 光谱变换方法的建模精度 R_c^2 最高可达 0.7, 但对应的预测精度 R_p^2 只有 0.33, 预测效果较差。所有光谱变换方法的建模精度 R_c^2 都在 0.65 以上 (除了 CR 仅为 0.59), 但只有 RL 光谱变换方法的估算精度 R_p^2 为 0.59, 其余均小于 0.5。综合来看, KS 样本集划分方法的最佳模型的光谱变换方法是 RL (R_c^2 为 0.66, R_p^2 为 0.59)。(3) 采用 SPXY 样本集划分方法, 最佳模型的光谱变换方法是 CR, 虽然其建模精度 R_c^2 仅为 0.58, 在所有方法中建模精度最小, 但其 R_p^2 为 0.49, 预测精度最高。其余光谱变换方法的估算精度 R_p^2 均小于 0.35, 甚至 FD 光谱变换方法的预测 R_p^2 出现了负值, 而该方法的建模精度 R_c^2 最高, 表明建模集和估算集的精度不具有—致性, 甚至可能完全相反。(4) 相同的光谱变换方法在不同的样本集划分方法下潜变量个数大致相同, 略有差异。例如, SNV 用 RS 和 KS 样本集划分的潜变量个数约为 7, 而用 SPXY 样本集划分为 6, 其余 CR 和 FD 方法也存在潜变量个数的差异。综合 PLSR 的建模和预测结果来看, KS 样本集划分方法结合 RL

光谱变换方法建立的土壤全磷含量偏最小二乘回归模型结果最优, R_c^2 为 0.66, R_p^2 为 0.59。

表 6 不同光谱变换方法和样本集划分方法的 PLSR 建模和估算结果

Table 6 PLSR calibration and prediction results of different spectral transformation methods and sample-set partitioning methods

光谱变换	RS			KS			SPXY		
	R_c^2	R_p^2	LVs	R_c^2	R_p^2	LVs	R_c^2	R_p^2	LVs
R	0.67	0.50	9.13	0.67	0.42	9	0.69	0.26	9
SNV	0.64	0.47	7.12	0.67	0.28	7	0.63	0.23	6
MSC	0.66	0.48	7.94	0.68	0.29	8	0.67	0.33	8
FD	0.63	0.46	4.38	0.70	0.33	6	0.71	-0.40	6
SD	0.68	0.44	5.01	0.68	0.46	5	0.68	0.27	5
RL	0.67	0.48	8.53	0.66	0.59	9	0.70	0.12	9
CR	0.61	0.48	4.93	0.59	0.35	4	0.58	0.49	4

从表 7 的 SVM 分析结果可得: (1) 采用 RS 样本集划分方法, FD 光谱变换方法的建模精度最高, R_c^2 为 0.91, 预测精度也最高 R_p^2 为 0.46。CR 和 SD 光谱变换方法效果次之, R_p^2 达 0.44 和 0.43, 对应的 R_c^2 均为 0.90, 其余光谱变换方法获得的建模结果较好, 预测结果均较低。(2) 采用 KS 样本集划分方法, FD 光谱变换方法的建模精度 R_c^2 最高可达 0.92, 对应的预测精度 R_p^2 有 0.82, 预测效果最好。其中 CR 的预测结果也较好, R_p^2 达 0.69, 其余方法的预测结果均低于 0.6。(3) 采用 SPXY 样本集划分方法, 最佳模型的光谱变换方法是 MSC, 虽然其建模精度 R_c^2 为 0.88, 在所有方法中建模精度中等, 但其 R_p^2 为 0.79, 预测精度最高。SD 和 FD 的建模精度 R_p^2 也达 0.78 和 0.76, 对应的建模精度 R_c^2 是 0.92 和 0.93, 三种方法精度相差较小。(4) 综合 SVM 方法的建模预测结果, KS 样本集划分方法结合 FD 光谱变换方法

表 7 不同光谱变换方法和样本集划分方法的 SVM 建模和估算结果

Table 7 SVM calibration and prediction results of different spectral transformation methods and sample-set partitioning methods

光谱变换	RS		KS		SPXY	
	R_c^2	R_p^2	R_c^2	R_p^2	R_c^2	R_p^2
R	0.84	0.20	0.80	0.41	0.83	0.46
SNV	0.89	0.37	0.89	0.55	0.84	0.49
MSC	0.89	0.37	0.89	0.56	0.88	0.79
FD	0.91	0.46	0.92	0.82	0.93	0.76
SD	0.90	0.43	0.90	0.45	0.92	0.78
RL	0.85	0.21	0.80	0.31	0.84	0.49
CR	0.90	0.44	0.90	0.69	0.90	0.71

建立的土壤全磷含量 SVM 模型结果最优, R_c^2 为 0.92, R_p^2 为 0.82。

综合两种方法的建模预测结果可以发现, 并非所有光谱变换方法都可以提高模型精度, 不同的样本集划分方法所适合的光谱变换方法也不一样, 尤其在 PLSR 模型中表现明显。采用 RS 方法划分样本集时 PLSR 的预测精度高于 SVM, 而采用 KS 和 SPXY 方法划分样本集, SVM 的预测精度整体高于 PLSR。采用 KS 方法划分样本集, PLSR 和 SVM 均能获得最优的预测结果, 但对应的光谱变换方法分别是 RL 和 FD。不同的样本集划分方法和光谱变换均能对建模预测结果产生较大的影响, 不同模型所适合的方法不同, 综合来看, 本研究区 KS 样本集划分方法结合 FD 光谱变换方法建立的土壤全磷含量 SVM 模型结果最优。

3 结 论

利用土壤反射光谱数据探讨了不同光谱变换结合不同样本集划分方法及预测模型在滨海土壤全磷含量估算中的可用性。本研究区土壤有机质与全磷相关性较小, 因此无法利用土壤有机质来获取全磷含量。从两种建模方法的结果来看, 在 PLSR 方法中, 1 000 次随机抽样结果表明有 45.70% 的可能性能得到合理且精度较高的土壤全磷含量模型, 利用 RS 方法划分样本集的结果在大部分情况下都优于 KS 和 SPXY 方法。而在 SVM 模型中, 1 000 次运算结果中只有 0.12% 的预测结果较好, 综合来说采用 SPXY 方法获得的结果最优, KS 次之, RS 结果最差。不同的样本集划分方法会对建模预测结果产生较大的影响, 不同的建模方法适合的样本集划分方法也有所差异, 在原始光谱数据的条件下, PLSR 模型适合 RS 方法划分样本集, SVM 模型适合采用 SPXY 方法划分样本。

经不同的光谱变换后, PLSR 和 SVM 模型采用 KS 方法划分样本集分别结合 RL 和 FD 变换的预测精度最高, R_p^2 为 0.59 和 0.82。因此, 预测本研究区土壤全磷含量的最佳模型是经过 FD 光谱变换, 采用 KS 方法划分样本集建立的 SVM 模型。综合两种建模结果看, 并非所有光谱变换方法都可以提高模型精度, 部分光谱变换后 PLSR 模型预测精度显著降低。其他样本集划分方法合适的光谱变换方法也不同, 对于 PLSR 模型, RS 结合原始光谱数据, SPXY 结合 CR 光谱变换获得的预测精度高; RS 结合 FD 光谱变换, SPXY 结合 MSC 光谱变换获得 SVM 模型预测精度高。总体而言, SVM 的建模效果优于 PLSR, 采用 RS 方法划分样本集, PLSR 的预测精度高于 SVM, 而采用 KS 和 SPXY 方法划分样本集, SVM 的预测精度整体高于 PLSR。

综上所述, 在实际应用中, 土壤样本采用不同样本集划分方法, 结合不同的光谱变换及建模方法会对土壤全磷含量的估算精度有较大影响, 考虑到区域性的差异, 普适性的建模方法及策略应该是下一步的研究方向。

References

- [1] ZHANG Jia-jia, GUO Xi, XIE Bi-yu, et al(张佳佳, 郭熙, 谢碧裕, 等). Journal of Jiangxi Agricultural University(江西农业大学学报), 2015, 37(3): 550.
- [2] XUE Li-hong, ZHOU Ding-hao, LI Ying, et al(薛利红, 周鼎浩, 李颖, 等). Acta Pedologica Sinica(土壤学报), 2014, 51(5): 993.
- [3] Xiong J F, Lin C, Ma R, et al. CATENA, 2019, 187: 104309.
- [4] WANG Li-wen, WEI Ya-xing(王莉雯, 卫亚星). Acta Ecologica Sinica(生态学报), 2016, 36(16): 5116.
- [5] Sarathjith M C, Das B S, Wani S P, et al. Current Science, 2016, 110(6): 1031.
- [6] CHEN Yi-yun, QI Tian-ci, HUANG Ying-jing, et al(陈奕云, 齐天赐, 黄颖菁, 等). Transactions of the Chinese Society for Agricultural Engineering(农业工程学报), 2017, 33(6): 107.
- [7] YU Hui-chun, LOU Nan, YIN Yong, et al(于慧春, 娄楠, 殷勇, 等). Food Science(食品科学), 2018, 39(16): 328.
- [8] MAO Bo-hui, SUN Hong, LIU Hao-jie, et al(毛博慧, 孙红, 刘豪杰, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2017, 48(S1): 160.
- [9] ZHANG Wen-long, ZENG Cong-sheng, GAO Deng-zhou, et al(章文龙, 曾从盛, 高灯州, 等). Acta Ecologica Sinica(生态学报), 2015, 35(24): 8085.
- [10] ZHANG Xiao-guang, JIANG Zi-xuan, KONG Fan-chang(张晓光, 姜子璇, 孔繁昌). Remote Sensing Technology and Application(遥感技术与应用), 2019, 34(4): 816.
- [11] JIN Hui-ning, ZHANG Xin-yue, LIU Huan-jun, et al(金慧凝, 张新乐, 刘焕军, 等). Acta Pedologica Sinica(土壤学报), 2016, 53(3): 627.
- [12] LIU Yan-fang, LU Yan-nian, GUO Long, et al(刘艳芳, 卢延年, 郭龙, 等). Acta Pedologica Sinica(土壤学报), 2016, 53(2): 332.
- [13] QI Hai-jun, LI Shao-wen, Karnieli A, et al(齐海军, 李绍稳, Karnieli A, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2018, 49(2): 166.
- [14] WANG Hai-yan, LI Jian-hui, YANG Feng-lei(汪海燕, 黎建辉, 杨风雷). Application Research of Computers(计算机应用研究), 2014, 31(5): 1281.
- [15] LU Long-mei, ZHANG Ping, LU Hong-liang, et al(陆龙妹, 张平, 卢宏亮, 等). Soils(土壤), 2019, 51(2): 374.

Estimation of Soil Total Phosphorus Content in Coastal Areas Based on Hyperspectral Reflectance

WEI Dan-ping, ZHENG Guang-hui*

School of Geographical Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, China

Abstract In recent decades, reflectance spectroscopy technology has developed rapidly and has been widely used in soil science, especially soil property estimation. It can greatly reduce the manpower and material resources consumed by traditional chemical measurement methods as an effective method to estimate total phosphorus content in soil. In this paper, 147 soil samples were collected from 30 sampling sites in coastal soil of Jiangsu Province, China. The spectral data and total phosphorus content of the soil were measured, respectively. Three different sample set partitioning methods were performed on the original spectral data and six different spectral transformation results, including Random Sampling (RS), Kennard-Stone (KS) and Sample Set Partitioning Based on Joint X-Y Distance Algorithm (SPXY). In order to compare and analyze the influence of three sample set partitioning methods on the accuracy of estimation results, partial least square regression(PLSR) and support vector machine (SVM) methods were used to establish the estimation models of total phosphorus content in the soil. The results are as follows: (1) Under the condition of original spectral data, the RS method can obtain better results and more stable model accuracy in most cases for PLSR, which is superior to KS and SPXY. In the SVM model, the result obtained by SPXY method is the best, KS is the second, RS is the worst. (2) The appropriate spectral transformation methods for different sample set partitioning methods are also different. Among the three sample set partitioning methods, the optimal spectral transformations of PLSR and SVM are respectively the reciprocal of logarithm and the first derivative (KS method), the original spectrum and the first derivative (RS method), the first derivative and multiple scattering correction (SPXY method). Using the KS method to divide the sample set, PLSR and SVM model can obtain the optimal prediction results. Not all spectral transformation methods can improve the model accuracy. The prediction accuracy of the PLSR model is significantly reduced after partial spectral

transformation. (3) Among all sample set partitioning methods, SVM has a better modeling effect than PLSR. Using the RS method to divide the sample set, the prediction accuracy of PLSR is higher than that of SVM. The results were reversed when KS and SPXY were used. According to the comprehensive results, the best estimation model for the study area was obtained using the KS sample set partitioning method, and the first derivative transformation method, combined with the SVM method, the R^2 of the prediction result was 0.82. This study shows that reflectance spectroscopy can effectively predict the total phosphorus content of the soil in coastal areas and have a certain guiding significance for the efficient and rapid inversion of soil phosphorus.

Keywords Total phosphorus; Reflection spectrum; Spectral transformation; Sample division method; Partial least squares regression; Support vector machine

(Received Dec. 29, 2020; accepted Jan. 28, 2021)

* Corresponding author

关于《光谱学与光谱分析》调整审稿费收费标准的通知

尊敬的《光谱学与光谱分析》广大作者、读者：我刊自 2018 年 7 月 1 日以后登记的稿件向投稿作者收取审稿费 200 元/篇，在您投稿之前，为免受经济损失，请您必须考虑：

1. 没有创新的一般性稿件，请您不要投稿。
2. 没有国家级基金资助的稿件，请您不要投稿。
3. 不是光谱专业的稿件，请您不要投稿。
4. 与其他文章重合率超过 10% 的稿件，请您不要投稿。

所投稿件经初审通过后，作者会收到缴纳审稿费的通知。请作者及时从我刊网站(<http://www.gpxygpx.com>)查询稿件是否处于交审稿费状态，在收到通知后，请及时缴纳审稿费；如在 10 天之内没有收到您的审稿费，被视为自动放弃，我刊不再受理。交费后我刊开具增值税电子普通发票，并传至作者提供的电子邮箱，作者可自行打印。

联系电话：010-62181070, 62182998

电子邮箱：chngpxygpx@vip.sina.com

感谢您多年来对《光谱学与光谱分析》的支持和厚爱！

《光谱学与光谱分析》期刊社

2018 年 6 月 30 日